Energetic Materials Frontiers 3 (2022) 177–186

Contents lists available at ScienceDirect

Energetic Materials Frontiers

journal homepage: www.keaipublishing.com/en/journals/energetic-materials-frontiers

Machine learning-guided property prediction of energetic materials: Recent advances, challenges, and perspectives

Xiao-lan Tian [a,b], Si-wei Song [a], Fang Chen [a], Xiu-juan Qi [b,*], Yi Wang [a,**], Qing-hua Zhang [a]

[a] Institute of Chemical Materials, China Academy of Engineering Physics (CAEP), Mianyang, 621999, China
[b] School of Material Science and Engineering, Southwest University of Science and Technology, Mianyang, 621010, China

ARTICLE INFO

Keywords:
Machine learning
Energetic materials
Property prediction

ABSTRACT

Predicting chemical properties is one of the most important applications of machine learning. In recent years, the prediction of the properties of energetic materials using machine learning has been receiving more attention. This review summarized recent advances in predicting energetic compounds' properties (e.g., density, detonation velocity, enthalpy of formation, sensitivity, the heat of the explosion, and decomposition temperature) using machine learning. Moreover, it presented general steps for applying machine learning to the prediction of practical chemical properties from the aspects of data, molecular representation, algorithms, and general accuracy. Additionally, it raised some controversies specific to machine learning in energetic materials and its possible development directions. Machine learning is expected to become a new power for driving the development of energetic materials soon.

1. Introduction

Chemical theory and calculations have been widely used to obtain the properties of energetic materials.[1–3] However, traditional molecular designs combined with theoretical calculations are often accompanied by problems, such as high computational cost or long estimation cycle, which are challenging due to the high-throughput computing requirements and experimental needs.[4] For example, density functional theory (DFT) requires expensive electronic structure calculations and a long cycle of geometric iterations, which results in high computational costs. Additionally, the consumption of chemical theory methods dramatically increases as the accuracy increases.[5] It is urgent to develop new methods to calculate the properties of energetic materials. Therefore, the emergence of machine learning enables the fast prediction of energetic materials' properties with accuracy approaching quantum chemical methods.[6]

The amount of data is rapidly expanding in the information age. Researchers can access information from massive data, acquire intrinsic mapping between molecular structures and properties, and finally predict the properties of unseen molecules.[7] This scientific research method is called the data-intensive or fourth scientific paradigm.[8] The rise in the application of the fourth paradigm to materials science implies that data analysis is increasingly used in material science research. Machine learning algorithms conform to the urgent need of the big-data era and have become powerful tools for data processing and model fitting. Among the different types of models, the supervised regression model, which is trained with continuously variable labels, is commonly used to determine energetic materials' properties.[9]

The key factors that affect the precision of regression algorithms in predicting molecular properties include the size and quality of data, featurization method, choice of algorithms, and hyperparameters tuning.[10] Typically, data collection is the first step for model training. The data resources mainly include experimental values from reported literature,[11–13] calculation results obtained using software programs (e.g., Gaussian and Vasp),[14] and formatted data from existing databases (e.g., CCDC and PubChem; Fig. 1a).[15,16] The second step is to choose methods to represent the chemical data, i.e., featurization. The most prevalent way for storing molecule structures is the Simplified Molecular Input Line Entry System (SMILES)—a method for representing molecular structures with ASCII strings.[17] Although SMILES can be regarded as a featurization method, it is usually leveraged as the starting point for constructing machine-readable features. Furthermore, traditional engineering-based featurization methods typically require extensive expertise in feature design and selection processes,[18] including Coulomb Matrix,

* Corresponding author.
** Corresponding author.
E-mail addresses: juanxiuqi@swust.edu.cn (X.-j. Qi), ywang0521@caep.cn (Y. Wang).

https://doi.org/10.1016/j.enmf.2022.07.005
Received 27 April 2022; Received in revised form 14 June 2022; Accepted 10 July 2022
Available online 18 August 2022

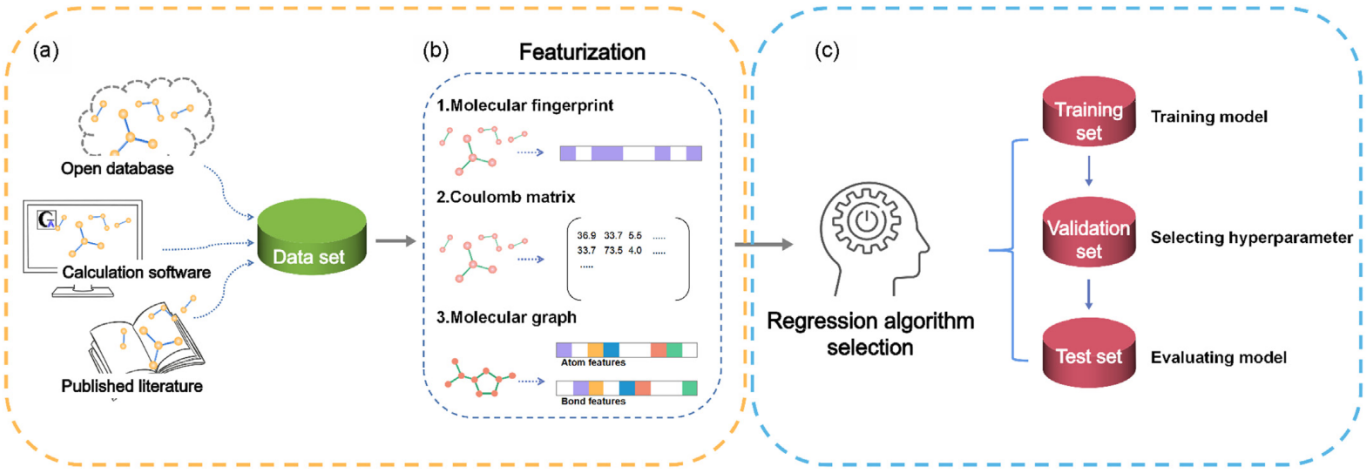

Fig. 1. Prediction flow charts using machine learning.

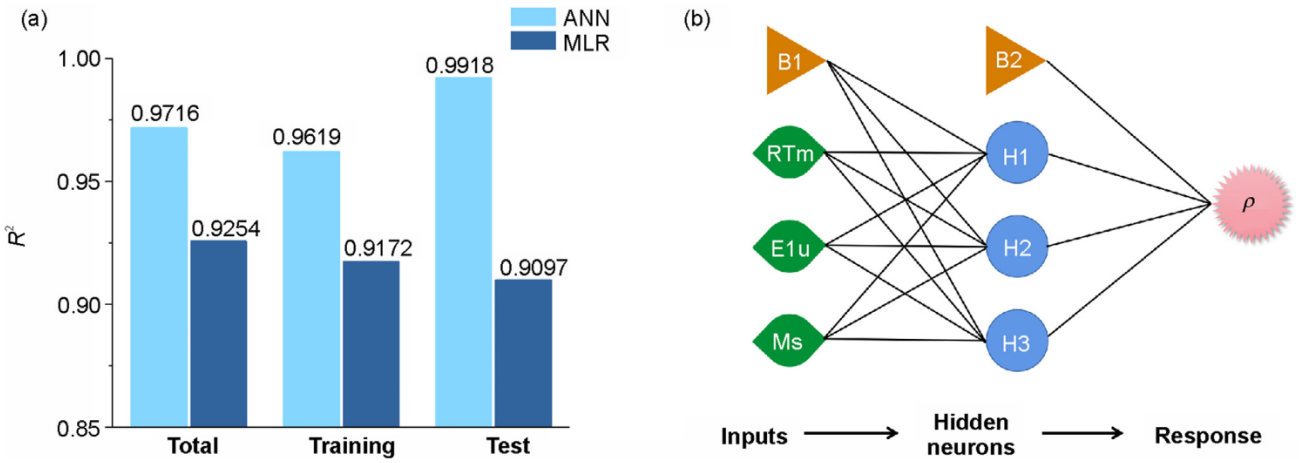

Fig. 2. (A) Histogram of error parameters of ANN and MLR, including the determination coefficients of total data set, training data set, and test data set from left to right along the x-axis; (b) ANN topological graph.[31].

extended-connectivity fingerprint, and custom descriptors.[19–22] With the maturity of artificial intelligence techniques, some scientists gradually utilize deep learning (e.g., graph neural networks to molecular graphs) to construct molecular fingerprints (Fig. 1b).[23] After determining the molecular representation method, it is necessary to identify the algorithms that are suitable for molecular property prediction, which is traditionally a regression task. There are many algorithms suitable for regression tasks, and the commonly used ones include linear regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), decision tree, kernel ridge regression, and neural network.[24,25] After training, the fitness of models can be evaluated using error metrics. The frequently used error metrics for the regression task include coefficient of determination (R^2), correlation coefficient(r), mean squared error (MSE), and mean absolute error (MAE).[26,27] Among them, r reflects the similarity between the measurements of two or more variables across a data set. R^2 represents the proportion of variance (of y) explained by the independent variables in the model and indicates the goodness of fit and, therefore, is a measure of the efficiency of the model in predicting unseen samples. MAE is a risk metric corresponding to the expected value of the absolute error loss. In addition, MSE is a risk metric corresponding to the expected value of the squared (i.e., quadratic) error or loss, which can be used as the loss function reflecting the algorithm's robustness. These analytical methods are derived from statistics. These methods allow for determining whether the selected structural attributes are beneficial in predicting properties and whether the constructed model can optimize the parameters. Generally, the data is split into three parts, i.e., training, validation, and test sets. The metrics on training, validation, and test sets help show the fitness of models, hyperparameter tuning, and the evaluation of the model performance on unseen data, respectively (Fig. 1c). These error metrics can be used to quantitatively compare the pros and cons of different models.

Machine learning-based predictions of the properties of energetic materials have recently been on the rise and achieved remarkable advancements in recent years. In the existing literature, some studies used the same o predict different properties,[28] some used different models to predict different properties,[29,30] and some used different models to predict the same properties. If we organized the studies according to predicted properties or machine learning models, the same study would be repeatedly mentioned in different sections, resulting in an overly verbose and loose-organized paper. Therefore, this review summarized recent advances in forecasting properties of energetic materials from two aspects, i.e., single property and multiple properties. This review compared the prediction accuracy and the advantages and disadvantages of each prediction method in detail. The purpose of this review is to reveal the research status of the energetic material property prediction based on machine learning, analyze the advantages of machine learning over traditional computing methods, and discover the existing problems in machine learning. Lastly, this review proposed the development

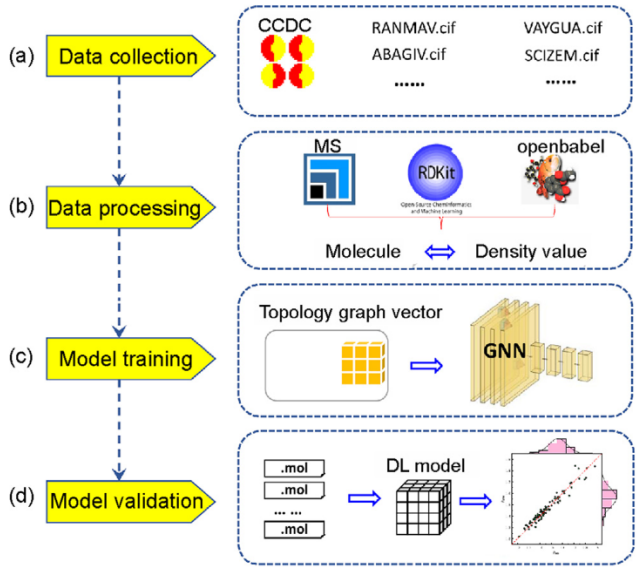

Fig. 3. Workflow chart and description of methods used in each step.[33] (a) Data collection; (b) Data processing; (c) Model training; (d) Model validation.

direction and prospect of machine learning in material property prediction in the future.

2. Single-property prediction

2.1. Density

Density is an important factor that affects the detonation performance of energetic materials. The detonation velocity and pressure of energetic molecules increase with increasing density. Traditional methods for predicting density mainly calculate the density using molecular or crystal volume, which can be accessed by theoretical or empirical methods, such as molecular dynamics or group additivity. Nevertheless, these methods always have some limitations. Molecular dynamics' accuracy heavily relies on the accuracy of the force field. If the applicable force field parameters are lacking, the prediction of crystal structure will have obvious deviation. The group additivity rules seldom consider molecular configuration and intermolecular interaction and thus are unable to distinguish the density of isomers. In addition, the influence of temperature and crystallographic form on density is neglected in group additivity rules, narrowing their applicability.[6] As the emerging forecasting method, machine learning can directly draw on the intrinsic relationship between structure and density while avoiding the effects of external conditions, thereby improving the computational efficiency and reducing the prediction error.

In 2018, Fathollahi et al.[31] extracted three molecular descriptors based on the optimized chemical structures of 26 selected energetic cocrystals. Their densities were predicted using an artificial neural network (ANN) and then the predicted results were compared with results from the multiple linear regression (MLR). The final calculation results are shown in Fig. 2a. According to this figure, the results predicted using ANN were larger than those of MLR, and the test precision was up to 0.9918. Fathollahi et al.[31] built a model consisting of an input layer with three neurons, a hidden layer with three neurons, and an output layer and determined that this model was ideal for valuation, as shown in Fig. 2b. This model could be used for screening the target chemical structure from the database to obtain the cocrystal with desired density. Moreover, Fathollahi et al.[32] predicted the decomposition temperature of energetic cocrystals using the ANN model again in 2018. The data set was divided into a training set containing 19 samples, a test set containing 6 samples, and a validation set containing 5 samples. The R^2 for ANN was 0.9784, which was higher than that of the MLR model (0.7438). Although a robust model used to predict property cannot be obtained using only 30 energetic cocrystals, the result reveals that ANN is an effective model for predicting decomposition temperature.

In 2021, Yang et al.[33] found that selecting molecular descriptors is time-consuming and relies on excellent professional knowledge. They decided to find a direct mapping relationship between density and molecular structure and take molecular topology as the sole known input. To this end, they filtered the crystallographic data of 2002 neutral nitro compounds from the CCDC database (Fig. 3a) and tested three algorithms, i.e., Random Forest (RF), Graph Neural Network (GNN), and Support Vector Machine (SVM), using inputs prepared collaboratively using Open Babel, Materials Studio, and RDKit library (Fig. 3b). Moreover, they compared the accuracy of predicted results of the DFT-QSPR, RF, SVM, and GNN models. The results indicated that the GNN model had the highest accuracy with R^2 up to 0.949, followed by the DFT-QSPR model with R^2 of 0.925. These results suggest that the GNN model can achieve high accuracy only using a molecular graph as the input and is significantly better than traditional machine learning models.

In 2021, Nguyen et al.[34] collected 10,521 molecules using data cleaning in CSD and trained density prediction models using four different machine learning algorithms, i.e., SVR, RF, PLSR, and MPNN—a GNN framework named by message passing neural network. When choosing data representation, they found that the SMILES was unsuitable for density valuation. Therefore, they chose three other different characterization methods (i.e., Extended 3D Fingerprint, 2D Molecular Descriptor Sets from RDKit, and Graph representation with Atom and Bond Descriptors from RDKit) for density prediction and evaluated the optimal computational scheme (Fig. 4).

Various molecular features in RDKit for different machine learning models were separately tested and ranked by their effects on density. Six feature descriptors (in order of VSA_EState8, SlogP_VSA5, TPSA, SMR_VSA5, MolLogP, and NO Count) were found to have more significant effects on density prediction. Moreover, MPNN with a learnable molecular representation had the highest computational accuracy (coefficient of determination up to 0.914) compared to manually-selected descriptors constructed by RDKit (Table 1), and the MPNN model has a relatively constant error less than 0.05 g cm^{-3} in most density intervals.

The aforementioned studies suggest that machine learning can effectively carve potential nonlinear relationships between molecular structure and density. Furthermore, it can alleviate the prediction error caused by models underfitting simplified parameters. Hence, machine learning can yield higher prediction accuracy than traditional empirical methods.[25] Additionally, Nguyen et al. and Yang et al. found that GNN models had a higher prediction accuracy. Though they adopted similar machine learning algorithms, e.g., GNN, the data sets used in the two studies were different. The different data distributions may influence the models and lead to differences in accuracy and generalization.

2.2. Detonation velocity

Detonation velocity is another important energetic property related to the structural properties and composition elements of energetic materials. As early as 2007, Ma et al.[35] combined neural networks with genetic algorithms for training detonation velocity prediction models. Accordingly, relevant parameters of explosives such as oxygen balance, molecular weight, and density were used as input. They compared the predicted values with experimental values of each compound, yielding errors within 7%. Although the calculation scheme they devised achieved rapid prediction with fewer relevant parameters, the accuracy cannot reach the valuation accuracy of the thermodynamic method based on the VLW equation of state.[36] Even so, the study carried out by Ma et al.[35] is instructive for the property prediction of energetic materials at that time.

In 2019, Chandrasekaran et al.[37] extracted data from over 65 compounds and developed an ANN-based detonation velocity prediction model that considered the effects of density, molecular structure,

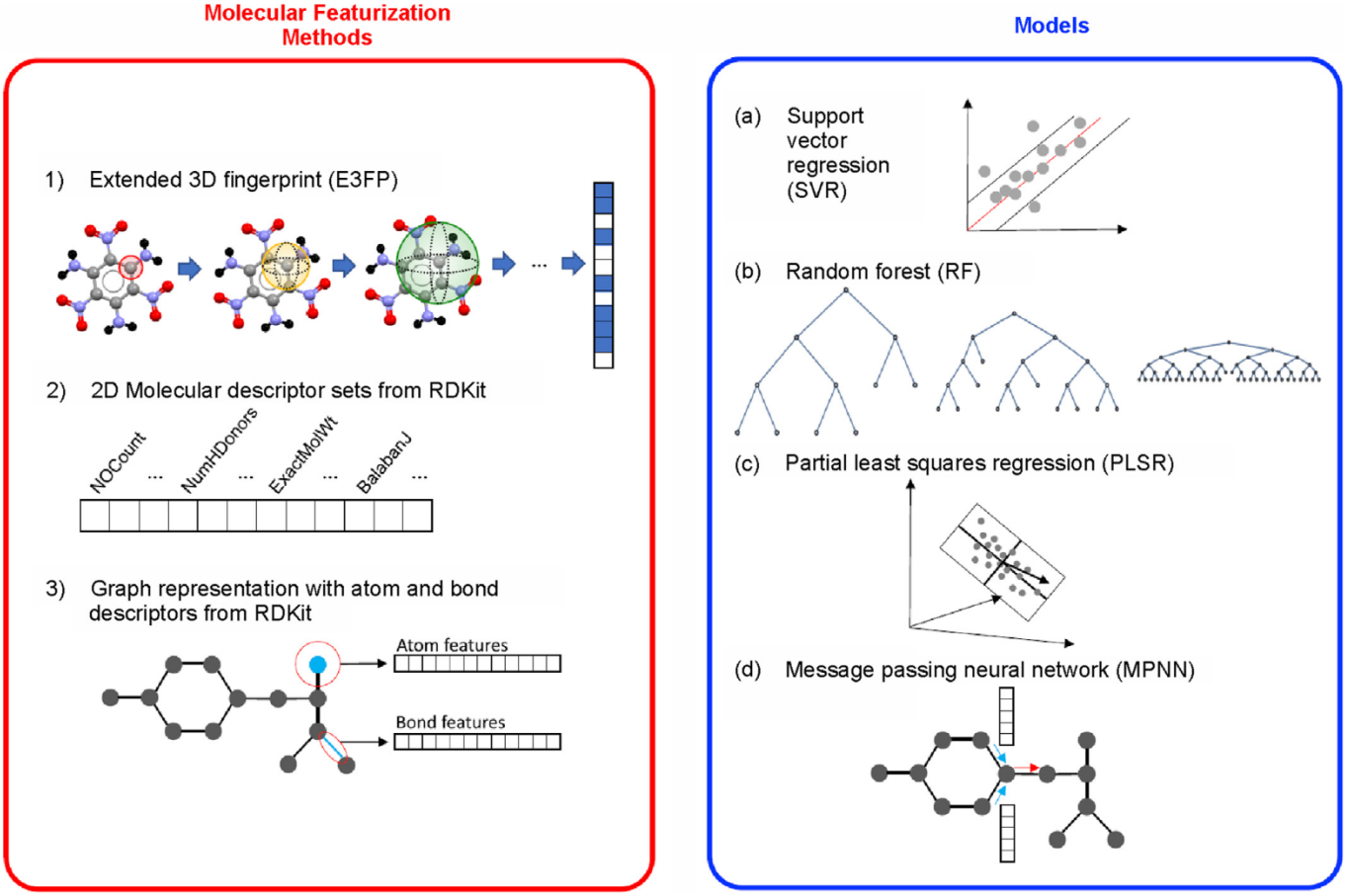

Fig. 4. Overview of molecular featurization methods (left) and models (right) adopted by Nguyen et al.[34].

Table 1
The R^2 and RMSE values of energetic material density predicted by combining different characterization methods with different models.[34].

Feature	Input information	Feature processing	Model	R^2	RMSE
E3FP	Atomic, 3D positions	precomputed	SVR	0.683	0.085
RDKit(molecular)	physicochemical/mathematical	precomputed	RF	0.878	0.053
RDKit(molecular)	physicochemical/mathematical	precomputed	PLSR	0.900	0.048
RDKit(atom/bond)	atomic/bond, molecule graph	learned	MPNN	0.914	0.044

chemical composition, and other characteristics. In this study, the ANN architecture consisted of one input layer, two hidden layers, and one output layer (Fig. 5a), and they fed enthalpy of formation, density, physical state, carbon, hydrogen, oxygen, and nitrogen atomic numbers into the input layer of neurons. The training data were composed of energetic materials with different characteristics and structures. The correlation coefficient between the training set and the test set of the model was up to 0.978–0.985, revealing that using ANN to predict the detonation velocity and taking highly correlated chemical information as input can achieve relatively high accuracy on a small data set.

2.3. Heat of formation

Regarding the thermodynamic properties of energetic compounds, the enthalpy of formation is an indispensable parameter in predicting the detonation performance of energetic materials. As early as 2004, Wang et al.[38] from Shenzhen University collected 58 aromatic polysonitro compounds and predicted the enthalpy of formation of polynitroaromatic compounds using a two-layer neural network of error back propagation (BP). They also discovered that multiple linear regression using the molecular map method had a significant effect and that most of the relative errors were controlled within 10%. The correlation coefficient of its regression equation was up to 0.9967. Meanwhile, they concluded that the molecular structural description code and network parameters greatly influence the prediction of the heat of formation.

In 2021, Mathieu[39] predicted electronic energy and vibration frequency using two deep-learning models (ANI-1X and ANI-1ccx), discovering that the enthalpy of formation for organic compounds can be calculated using the standard atom equivalent scheme along with predicted energy and frequency. He demonstrated that the accuracy of deep learning models was comparable to that of costly density functional theory calculation, especially for general CHON compounds. However, considering the decrease in model accuracy for energetic compounds, he concluded that current models were not ready to be applied to energetic compounds (Fig. 5c). This result may be due to the lack of explosophore-containing compounds in the GDB-11 database, which was used to fit the current ANI model. Further refitting models on data containing more energetic compounds may improve the models' applicability to energetic compounds. Meanwhile, there is an urgent need to develop extensive ab initio databases, including energetic molecules.

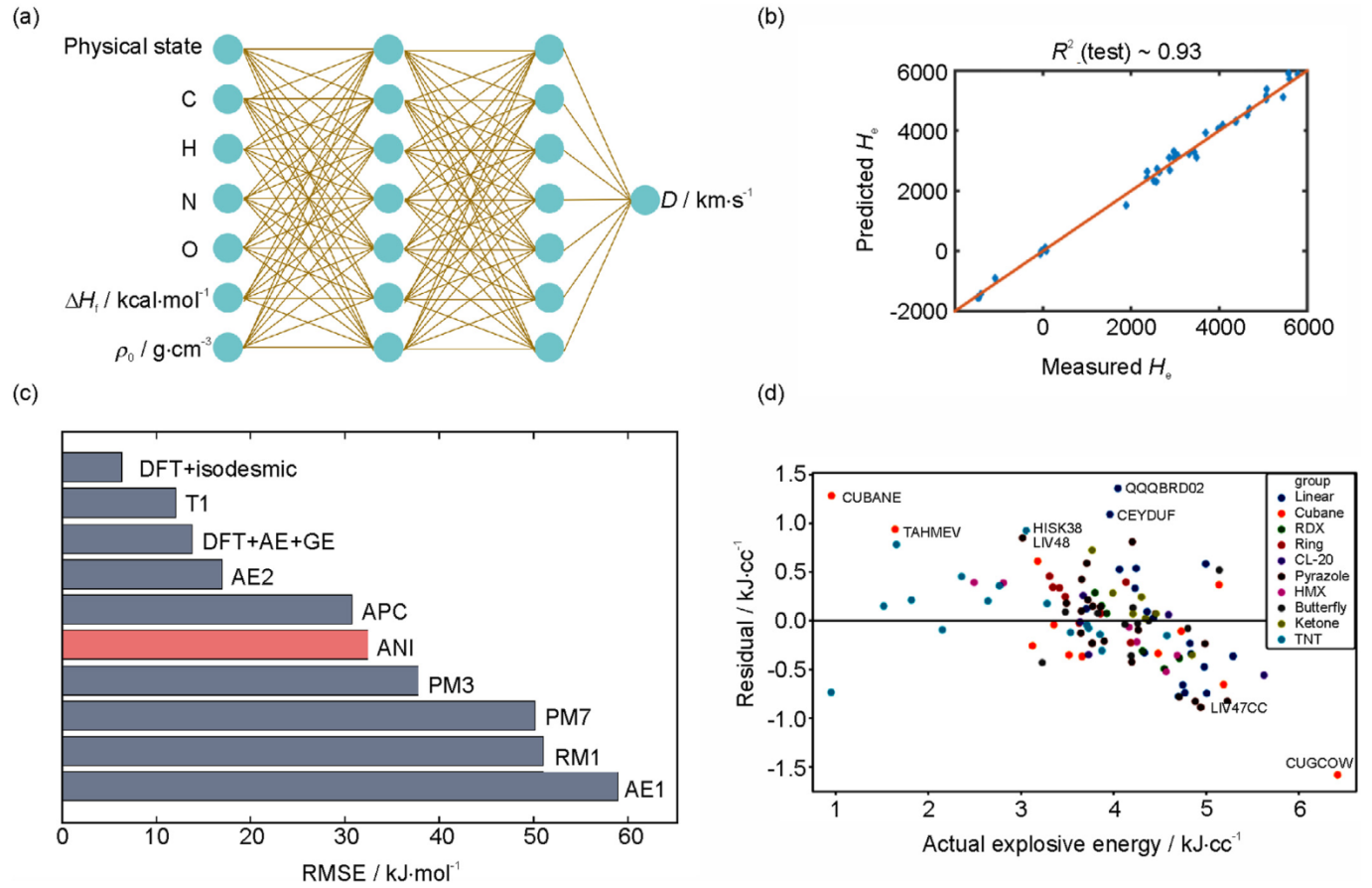

Fig. 5. (A) ANN architecture in Chandrasekaran's study.[37] (b) Kang compared the predicted values of ML with measured values collected (the red line indicates exact prediction, i.e., the predicted value equals the measured value).[41] (c) Average performance of ANI compared to alternative schemes in HEDM-45 (after Mathieu[39]). (d) Elton prepared the plot of error distributions for leave-one-out cross-validation with the Kernel ridge regression and sum over bonds.[22].

2.4. Heat of explosion

The heat of explosion is closely related to the detonation velocity and detonation pressure of energetic materials. Based on training data set composed of 41 energetic compounds, Kang et al.[41] trained machine learning models for predicting the heat of explosion. They found that oxygen balance and atomic average cohesive energy were the two most influential descriptors for predicting the heat of explosion. ML algorithms were applied for training models, including linear regression, logistic regression, random forest, and support vector machine. The error analysis showed that the lowest mean absolute error (MAE) for the models trained using RF was 142.12 kJ kg^{-1} and that the corresponding R^2 was up to 0.93 (Fig. 5b). Using the machine learning models, 2732 candidate molecules in the PubChem and ICSD databases were screened using the heat of explosion of TNT as the standard.

2.5. Sensitivity

Sensitivity marks the difficult degree of energetic materials to burn or explode under the impact, friction, flame, and electrostatic stimuli. Higher sensitivity is associated with lower safety of energetic materials. Therefore, it is necessary to obtain sensitivity in research on energetic molecules. Keshavarz et al.[40] conducted in-depth studies of the sensitivity prediction of energetic materials using artificial neural networks. In their study, they adopted 291 energetic compounds containing C, H, O, and N elements and divided the data set into training and test sets. ANN models used descriptors consisting of the number of aromatic characters, heteroaromatic characters, N–NO_2 bonds, and the C, H, O, and N elements as input. The reliability of models was further tested using 14 different explosives with various chemical structures. The predictions using the constructed ANN models yielded root means square errors of the training and test data of 41 cm and 56 cm, respectively.

3. Multi-property prediction

It is established that the comprehensive analysis of an energetic compound requires the prediction of its different properties. Thus, predicting multiple properties by “one run” is more attractive because this prediction mode is more efficient and friendly. In recent years, some scientists have conducted related research on multi-property prediction under a unified model architecture (e.g., using the same input and algorithm for predicting different properties). Meanwhile, some researchers are engaged in evaluating various combinations of featurization methods and ML algorithms, aiming to find the optimal combination for the simultaneous prediction of multiple attributes.

In 2018, Elton et al.[22] investigated machine learning techniques for predicting multiple properties of energetic materials. They worked on a data set containing 109 molecules with 10 different chemical structures. They predicted the properties of energic compound (e.g., enthalpy of formation, density, detonation velocity, detonation pressure, and detonation energy) using various featurization methods (e.g., bag of bonds, coulomb matrices, sum over bonds, and custom descriptor set) and ML algorithms (e.g., ridge regression, kernel ridge regression, random forest, k-nearest neighbors, and support vector regression). The results revealed that the characteristic values, such as oxygen balance, nitrogen-carbon ratio, sum over bonds, and the number of functional groups were essential parameters affecting the computation accuracy. Among them, sum over bonds was the most favorable feature descriptor for improving

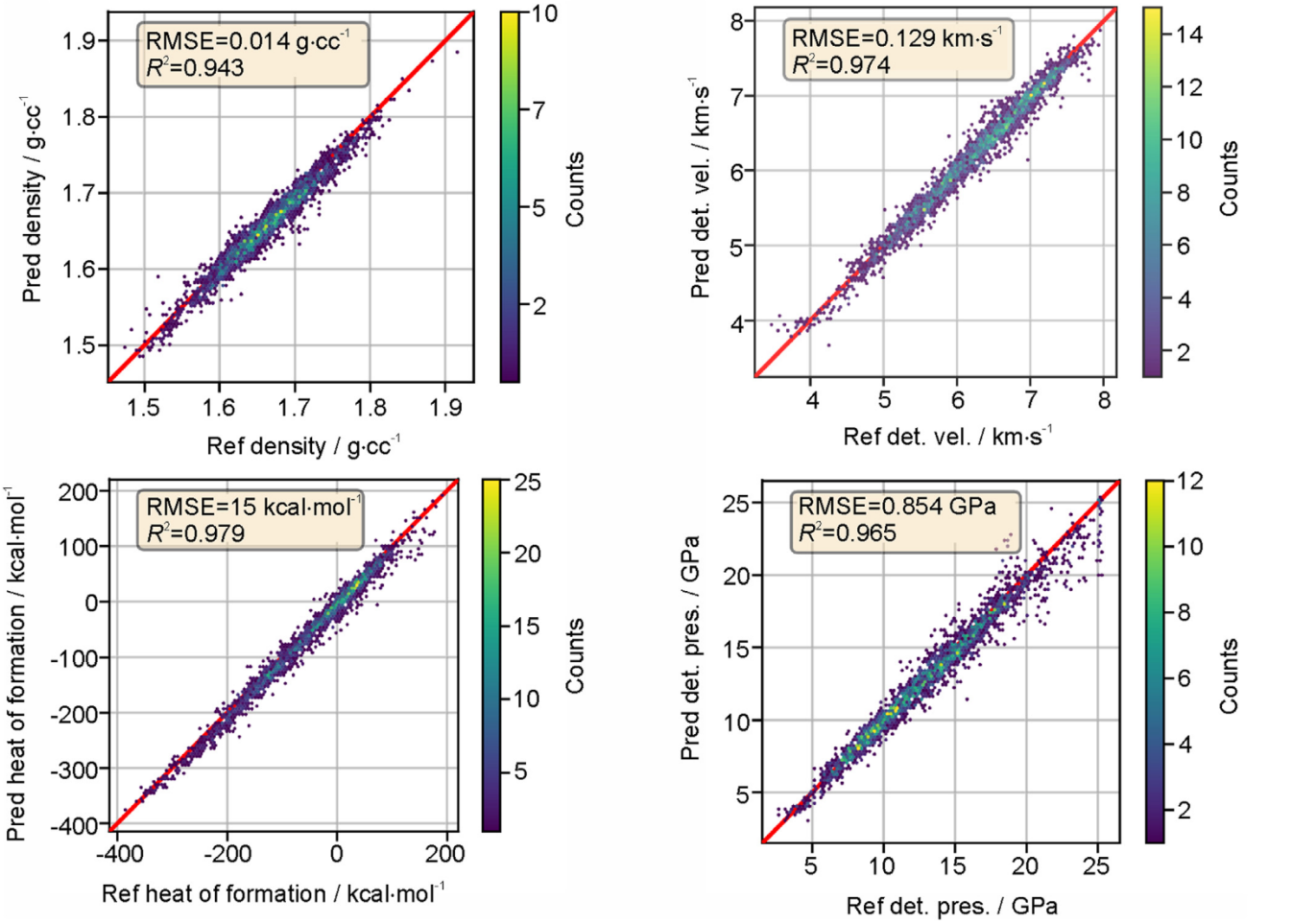

Fig. 6. Parity plots of prediction values (CNN) *vs.* actual values of different properties on the test set (the red line indicates exact prediction, i.e., the predicted value equals the actual value).[14].

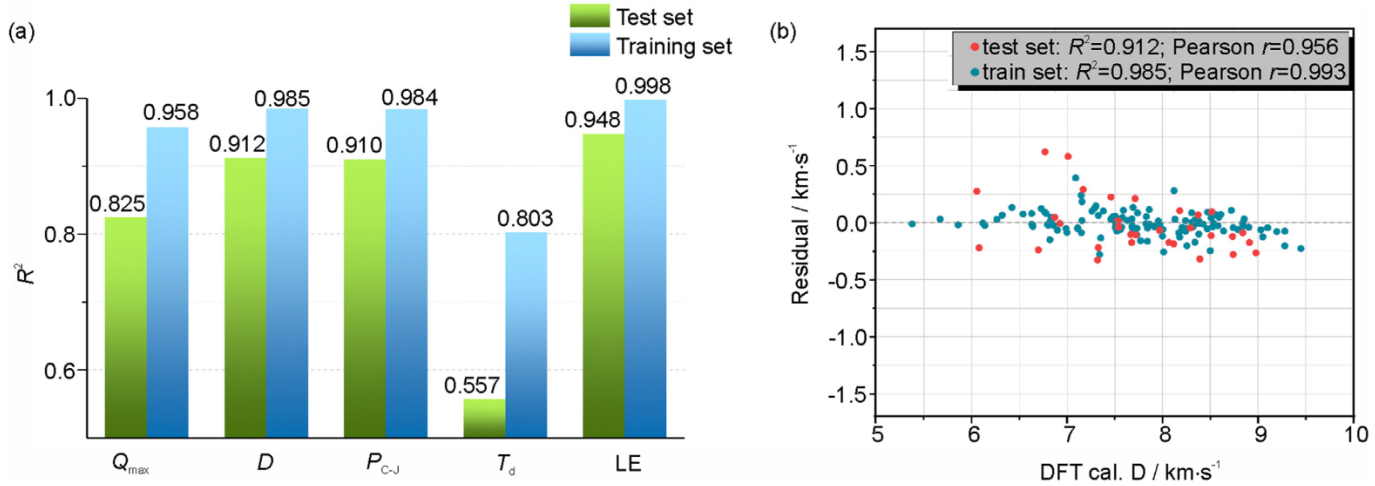

Fig. 7. (A) Diagram showing coefficient of determination distributions of the heat of explosion, detonation velocity, detonation pressure, decomposition temperature, and lattice energy prediction using XGBoost from left to right. (b) XGBoost predicted residuals for detonation velocity.[44].

prediction accuracy, and kernel ridge regression was the optimal algorithm (Fig. 5d). They proposed that feature selection was more important than model selection when making property predictions on small datasets and that sample diversity was vital for achieving better accuracy in the case of less available data. Elton et al. focused on a small data set that contributed to fields with few available research results or limited public databases (such as energetic materials[42]). The study carried out by Elton et al. can help researchers obtain performance prediction values with less

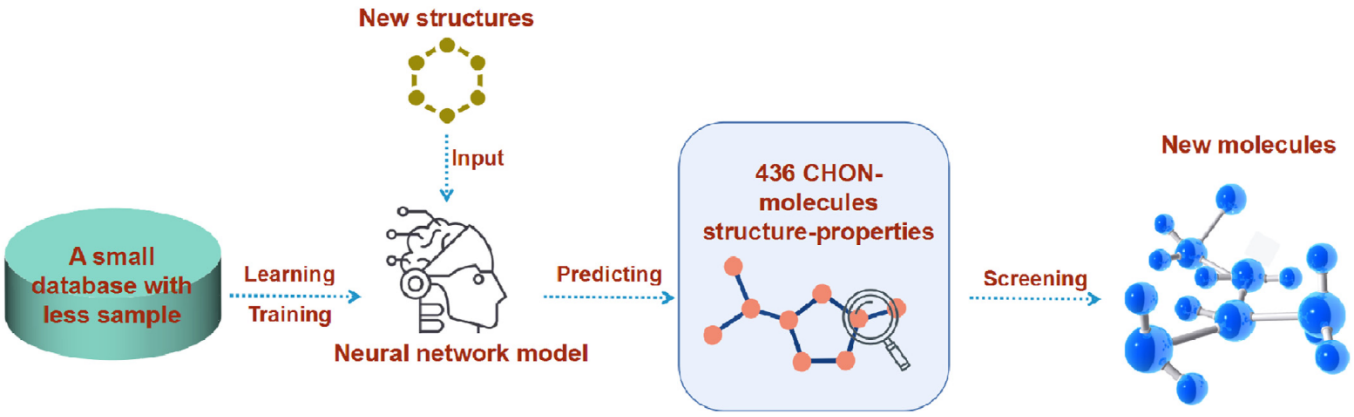

Fig. 8. Flow chart showing the machine learning for performance prediction and molecular screening.

Table 2
MAE and *r* for the property prediction results obtained using the neural network model.[45].

Properties	Unit	*r*	MAE
ρ	g cm^{-3}	0.9860	0.0259
D	km s^{-1}	0.9279	0.3456
P	GPa	0.9664	1.4933

error in the case of a small amount of available information.

In 2020, Casey et al.[14] predicted the detonation temperature, detonation velocity, detonation pressure, heat of formation, and density of energetic materials using 3D convolutional neural networks (CNN). By screening molecules with energetic properties using oxygen balance as a parameter, they identified more than 20,000 molecules from the GDB database. Furthermore, they calculated electronic and energetic properties using Gaussian and Cheetah while utilizing the results as training data for a convolutional neural network. R^2 of the models they used exceeded 0.9, reaching the ideal error range (Fig. 6). These models achieved impressive accuracy, possibly proving that high-quality data and model architecture with adequate complexity can pave the way for high accuracy prediction. However, owing to the scarcity of energetic compounds in the GDB database, the applicability of these models for energetic compounds is yet to be verified.

In 2021, Chen et al.[43] demonstrated two new feature extraction methods, namely heat contribution spatial matrix and volume occupation spatial matrix. They built the machine learning model with a data set of 451 energetic molecules, with one-fifth of the molecules used as a testing set and the rest serving as the training data. Furthermore, 56 candidate energetic molecules with reasonable chemical structures and excellent detonation properties were screened. The spatial matrix descriptors used in the machine learning model considered the heat contribution and volume occupation from monoatomic (1-body) and interatomic (2-body) aspects. The machine learning model built by Chen et al.[43] could become more powerful if it focused on more physicochemical features and a higher level of many-body aspect.

Rein et al.[28] predicted the impact sensitivity and decomposition temperature of high energetic materials using a multiple linear regression model (MLR). From various literature, they collected experimental crystal structures and sensitivity properties of more than 400 energetic materials, including materials that are mainly developed nowadays, thus obtaining sufficient data for MLR modeling. Subsequently, they successfully predicted the properties of nitrogen-rich tetrazoles and azides using used the established MLR.

Sifain et al.[29] proposed a method for predicting molecules' melting and boiling points using ridge regression (UPPER-RR) and gradient boosting (UPPER-GB). They found that the prediction accuracy of enthalpy and entropy was not improved using the UPPER-GB and UPPER-RR. Meanwhile, they discovered that the predicted results of ARL-UPPER were more accurate than those of the first two methods. They split all data into training and test sets in a ratio of 9:1 and compared the prediction error of ARL-UPPER with that of Extreme Gradient Boosting (XGBoost), which was combined with 1024-bit Morgan, 1024-bit Avalon fingerprints, and 166 MACCS. The results indicated that the ARL-UPPER had significantly higher accuracy than other methods. The cross-validation RMSE of boiling and melting points predicted using ARL-UPPER was 20K and 36K, respectively, making ARL-UPPER framework a potential materials screening tool.

Xie et al.[30] developed a Property-Oriented Adaptive Design Framework (PADF) to rapidly design energetic molecules with excellent performance. They selected 88 compounds as training data and predicted the heat of explosion and heat of formation using used the PADF. In their study, they considered four feature descriptors (i.e., Sum Over Bonds (SOB), Extended Connectivity Fingerprint (ECFP), E-State Fingerprint (E-state), and Custom Descriptor Set (CDS)) and six ML algorithms (i.e., Least Absolute Shrinkage and Selection Operator (LASSO), Linear Regression (Lin), Support Vector Regression with a Linear Kernel (SVR.lin) and a Radial Basis Kernel (SVR.rbf), Kernel Ridge Regression (KRR), and a Gaussian Process Regression (GPR) model). In the PADF, they also considered five optimizers and selected the best performing one. Through experiments, they finally concluded that SVR.lin/Trade-off combined with E-state + SOB was the best model for calculating the heat of formation, with the R^2 and MAE of the test set up to 0.93 and 61.7, respectively. Moreover, KRR/KG coupled with CDS + E-state + SOB was the best for predicting the heat of explosion.

Huang et al.[44] applied machine learning to 153 high-energy density materials. All data were obtained through high-throughput crystal-level quantum mechanics calculations and were divided into training data and test data at a ratio of 4:1. They conducted the model training using KRR, RF, XGBoost regression tree, adaptive boosting (AdaBoost) regressor, and multilayer perceptron (MLP) and compared the predicted results with the experimental values. The comparison results indicated that the XGBoost model yielded excellent performance in the prediction of nearly all properties. Huang et al. collected 109 experimental values of decomposition temperature (T_d) and 612 calculated data on detonation velocity (*D*), detonation pressure ($P_{C\text{-}J}$), heat of explosion (Q_{max}), and lattice energy (LE) as well as 203 experimental values to validate the model's predictions. As shown in Fig. 7, the determination coefficients of all properties (except for decomposition temperature) on the test set were up to more than 0.8.

Hou et al.[45] developed a neural network model that can be used to effectively predict the properties of energetic molecules, including detonation velocity, pressure, and density. They placed many typical explosives (e.g., TNT, CL-20, HMX, and RDX) in their initial dataset and screened candidate molecules from 436 molecules using NN models (Fig. 8). The properties of compounds in the training data were evaluated

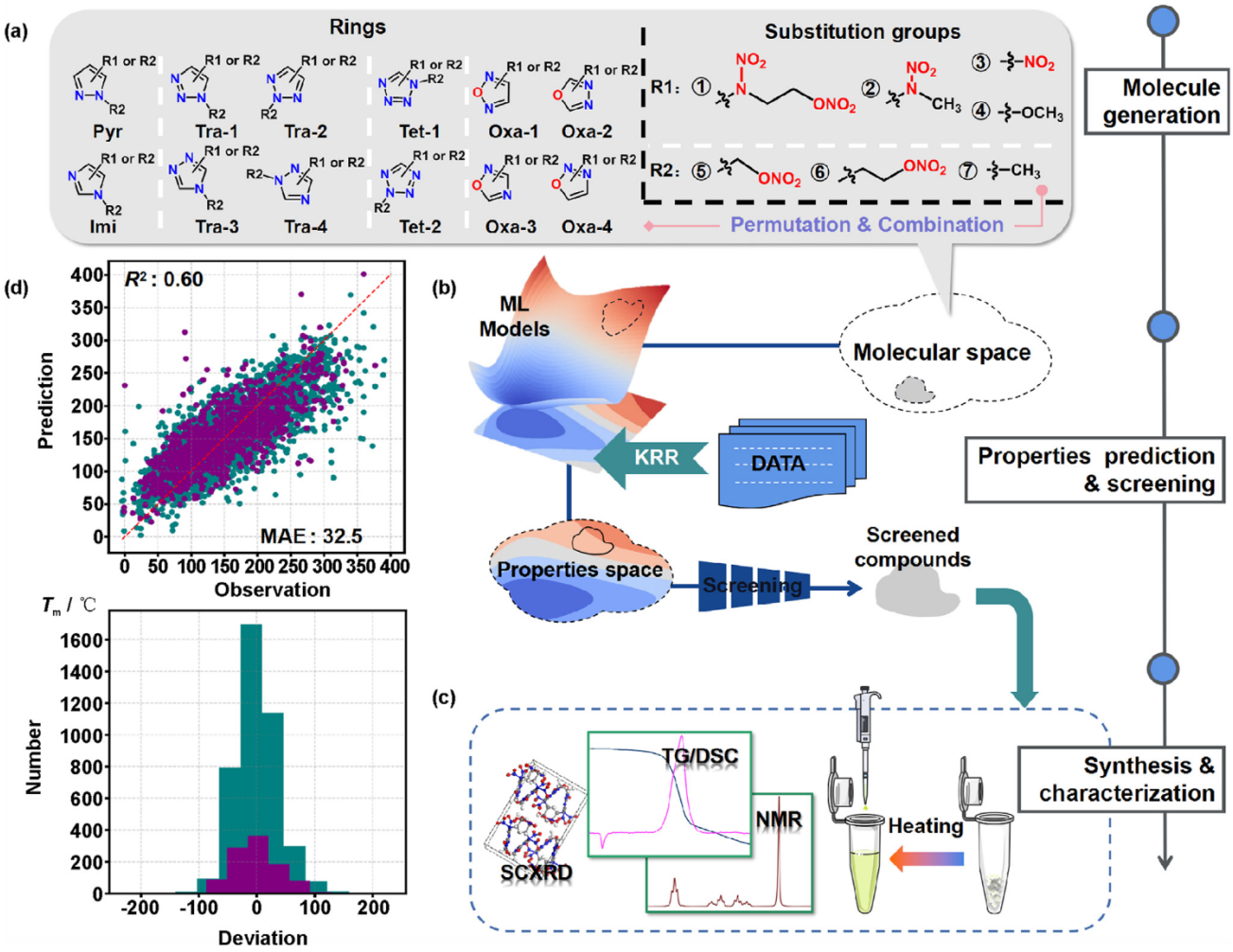

Fig. 9. Machine learning-assisted the overall framework of HTVS.[47] (a) Molecular fragments used to construct molecular libraries. (b) Schematic diagram of machine learning-assisted property prediction and molecular screening of targeted compounds. (c) Synthesis and characterization. (d) Predicted values vs. observed values and error distribution for the melting point.

using theoretical methods, such as the density equation developed by Politzer et al. and the Kamlet-Jacobs equation for detonation properties.

Hou's machine learning model only took 0.038 s to predict the performance of three kinds of detonation for all molecules, thereby demonstrating the high-throughput characteristics of machine learning. Table 2 shows that the predicted values of the detonation performance were significantly reduced by applying the model developed by Hou et al. indicating that their neural network model can yield excellent performance and can accurately predict the detonation characteristics of energetic molecules.

In addition to the above studies, this review combined machine learning with high-throughput virtual screening (HTVS) and experiments to advance the exploration of energetic materials.[46,47] In 2021, Song et al.[47] used the machine learning model to assist the HTVS system in high-throughput molecular screening. As a result, they quickly selected 136 molecules with a satisfactory performance from more than 3800 molecules and finally obtained eight new melt casting materials through experimental research. The specific experimental process is shown in Fig. 9a–c. Furthermore, they manually collected more than 1000 pieces of data from existing literature to construct the database required for the machine learning model. Molecular structures were represented by feature vectors using E-state fingerprints and custom descriptors. They applied the kernel ridge regression algorithm to performance prediction, yielding high prediction accuracy of density and detonation properties. However, the predicted results of the melting point © and T_d did not fulfill the expectations. By comparison with feature weights from the RF model, they concluded that the descriptors related to intermolecular interactions, such as Topological Polar Surface Area (TPSA), Number of Rings (*nR*), and Min/Max Partial Charge, had a crucial influence on the melting point calculations. Therefore, the descriptors customized by Song et al.[47] are instrumental in predicting melting points.

Since density, enthalpy of formation, detonation velocity, and detonation pressure of energetic compounds are generally concerned, this review summarized metrics of related models, as shown in Table 3. Through continuous endeavor, predicting the density of energetic compounds using machine learning has significantly developed and can achieve high prediction accuracy compared to other properties. The coefficient of determination for density in these studies has exceeded 0.9, with the MAE not higher than 0.06 g cm^{-3}. Meanwhile, the high prediction accuracy of detonation velocity has also been achieved, with coefficients of determination all higher than 0.8 in these studies. Casey's 3D CNN models have prediction accuracy of detonation velocity of up to 0.991 on the training set and up to 0.974 on the test set (Table 3). Additionally, these 3D CNN models enjoy advantages in predicting other properties such as density and enthalpy of formation. The synergistic contribution of larger data, a more complex model, and reasonable featurization synergistic yields higher

Table 3
Key information from the above literature, including the amount of data, machine learning algorithm, chemical information tools, and some properties prediction accuracy (the best machine learning method is in bold).

Author	Data pieces	Algorithms	Featurization	ρ		ΔH_f		D_v		P	
				R^2	MAE /g·cm^{-3}	R^2	MAE /J·mol^{-1}	R^2	MAE /km·s^{-1}	R^2	MAE /GPa
Fathollahi[31]	26	**ANN**, MLR	Gaussian 09, Dragon	0.992 (test)	–	–	–	–	–	–	–
Yang[33]	2002	SVM, RF, **GNN**	Materials Studio, RDKit, OpenBabel	0.949 (test)	0.040 (test)	–	–	–	–	–	–
Nguyen[34]	10521	SVR, PF, PLSR, **MPNN**	RDKit, E3FP	0.914	–	–	–	–	–	–	–
Chandrasekaran[37]	65	**ANN**	Custom descriptor set	–	–	–	–	0.985[r] (test)	–	–	–
Elton[22]	109	**KRR**, Ridge, SVR, RF,KNN	Custom descriptor set, Coulomb matrix, RDKit	0.74[r]	0.060 (test)	0.94[r]	71.41 (test)	–	–	–	2.760 (test)
Casey[14]	26265	**3D CNNs**	Grid data for electron charge density and electrostatic potential	0.943 (test)	0.011 (test)	0.979 (test)	47.09 (test)	0.974 (test)	0.096 (test)	0.965 (test)	0.584 (test)
Chen[43]	451	LASSO, **KRR**, BRR, SVR, RFR, KNN	VOM, HCM	–	0.035	–	40.44	–	–	–	–
Xie[30]	88	LASSO, Lin, SVR.lin, SVR.rbf, KRR, GPR	SOB, ECFP, E-state, Custom descriptor set			0.93 (test)	258.28 (test)				
Huang[44]	153	**XGBoost,** AdaBoost, RF, MLP, KRR	Custom descriptor set	–	–	–	–	0.912 (test)	–	0.910 (test)	–
Hou[45]	436	**LM**	Coulomb matrix	0.986 [r]	0.026	–	–	0.928[r]	0.346	0.966 [r]	1.493
Song[47]	1000	**KRR**	E-state fingerprint, custom descriptors	0.930	0.042	–	–	0.830	0.240	0.820	2.379

accuracy. However, the accuracy improvement for Casey's models is at the expense of a relatively high cost in calculating grid data of electron charge density and electrostatic potential. Even so, this study has paved the way for models with high accuracy.

4. Conclusions and outlooks

The above research shows that machine learning has made significant progress in predicting the properties of energetic materials. Starting with only molecular structures, machine learning can yield high overall prediction accuracy for density, detonation properties, and enthalpy of formation at an extremely low computational expense. However, machine learning cannot yet yield satisfactory prediction accuracy for decomposition, melting temperature, and sensitivities, which are difficult to precisely predict using methods based on electronic structures or molecular dynamics. The reasons include: (1) the noise in the data caused by experimental conditions, (2) intrinsic multiscale characteristics from quantum scale to continuum mechanics, and (3) the complex thermochemical coupling process behind these properties. It can be foreseeable that these problems are difficult to overcome within a short period.

Although it is difficult to precisely predict the properties suffering difficult prediction while ensuring the model's generalization ability in a short time, there are many ways worth looking into to improve the precision. Possible methods include more complicated model architectures trained on large data (e.g., GNN) and more complete and physically meaningful molecular representation. Additionally, feature engineering and transfer learning are viable solutions to improve model performance. However, these methods usually mean an increase in model complexity and require a large amount of data to avoid over-fitting. Accordingly, the lack of data is another struggle that limits the application of machine learning in the field of energetic materials due to military applications and hazards. Hence, there is an urgent need to establish a standard database of energetic materials, which may require the collaboration of the entire energetic materials community.

At present, the application of machine learning techniques in the field of energetic materials is still at its initial stage. Besides property prediction, machine learning can also be applied in inverse molecular design, automatic synthetic route planning, and machine learning force field. These cutting-edge machine learning applications will promote the development of energetic materials from design, experiment, and simulation aspects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This review was supported by the National Natural Science Foundation of China (Grant Nos. 22075259, 22175157).

References

1. Li G, Zhang C. Review of the molecular and crystal correlations on sensitivities of energetic materials. *J Hazard Mater*. 2020;398, 122910.
2. Guo ST, Liu J, Qian W, Zhu WH, Zhang CY. A review of quantum chemical methods for treating energetic molecules. *Energ Mater Front*. 2021;2(4):292–305.
3. Chen F, Song SW, Wang Y, Liu Yc,Zhang QH. Effects of alkyl chains on the physicochemical properties of nitroguanidine derivatives. *Energ Mater Front*. 2020; 1(3-4):157–164.
4. Wang Y, Liu YJ, Song SW, et al. Accelerating the discovery of insensitive high-energydensity materials by a materials genome approach. *Nat Commun*. 2018;9:2444.
5. Graser J, Kauwe SK, Sparks TD. Machine learning and energy minimization approaches for crystal structure predictions: a review and new horizons. *Chem Mater*. 2018;30(11):3601–3612.
6. Wang LL, Xiong Y, Xie WY, Niu LL, Zhang CY. Review of crystal density prediction methods for energetic materials. *Chin J Energetic Mater*. 2020;28(1):1–12.
7. Deng QQ, Hu J, Wang LY, et al. Probing impact of molecular structure on bulk modulus and impact sensitivity of energetic materials by machine learning methods. *Chemometr Intell Lab Syst*. 2021;215, 104331.
8. Agrawal A, Choudhary A. Perspective: materials informatics and big data: realization of the 'fourth paradigm' of science in materials science. *Apl Mater*. 2016;4(5), 053208.
9. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature*. 2018;559:547–555.
10. Wang HS, Ji YJ, Li YY. Simulation and design of energy materials accelerated by machine learning. *WIREs Comput Mol Sci*. 2019;10(1), e1421.

11. Morrill JA, Byrd EFC. Development of quantitative structure property relationships for predicting the melting point of energetic materials. *J Mol Graphics Modell*. 2015; 62:190–201.
12. Wang R, Jiang JC, Pan Y, Cao YH, Cui Y. Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices. *J Hazard Mater*. 2009;166(1):155–186.
13. Balakrishnan S, VanGessel FG, Boukouvalas Z, Barnes BC, Fuge MD, Chung PW. Locally optimizable joint embedding framework to design nitrogen-rich molecules that are similar but improved. *Mol Inf*. 2021;40(7), 2100011.
14. Casey AD, Son SF, Bilionis I, Barnes BC. Prediction of energetic material properties from electronic structure using 3D convolutional neural networks. *J Chem Inf Model*. 2020;60:4457–4473.
15. Fortunato ME, Coley CW, Barnes BC, Jensen KF. Machine learned prediction of reaction template applicability for data-driven retrosynthetic predictions of energetic materials. *AIP Conf Proc*. 2020;2772(1), 070014.
16. Barnes BC. Deep learning for energetic material detonation performance. *AIP Conf Proc*. 2020;2272(1), 070002.
17. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci*. 1988;28(1):31–36.
18. Guyon I, Elisseef A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003;3:1157–1182.
19. Montavon G, Rupp M, Gobre V, et al. Machine learning of molecular electronic properties in chemical compound space. *New J Phys*. 2013;15(9):501–509.
20. Le T, Winter R, Noé F, Clevert D. Neuraldecipher - reverse-engineering extended-connectivity fingerprints (ECFPs) to their molecular structures. *Chem Sci*. 2020;11: 10378–10389.
21. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*. 2010;50: 742–754.
22. Elton DC, Boukouvalas Z, Butrico MS, Fuge MD, Chung PW. Applying machine learning techniques to predict the properties of energetic materials. *Sci Rep*. 2018;8: 9059.
23. Ma HH, Bian YT, Rong Y, et al. Cross-dependent graph neural networks for molecular property prediction. *Bioinformatics*. 2022;38(7):2003–2009.
24. Mariappan A, Choi H. The application of energetic materials genome approach for development of the solid propellants through the space debris recycling at the space platform. In: *AIAA Propulsion and Energy 2020 Forum*. 2020:3898.
25. Juan YF, Dai YB, Yang Y, Zhang J. Accelerating materials discovery using machine learning[J]. *J Mater Sci Technol*. 2021;79:178–190.
26. Gubaev K, Podryabinkin EV, Shapeev AV. Machine learning of molecular properties: locality and active learning. *J Chem Phys*. 2018;148(24), 241727.
27. Chen C, Zuo YX, Ye WK, Li Xg,Deng Z, Ong SP. A critical review of machine learning of energy materials. *Adv Energy Mater*. 2020;10(8), 1903242.
28. Rein J, Meinhardt JM, Hofstra Wahlman JL, Sigman MS, Lin S. *An Explosophore-Based Approach towards the Prediction of Energetic Material Sensitivity Properties. ChemRxiv*. Cambridge: Cambridge Open Engage; 2021.
29. Sifain AE, Rice BM, Yalkowsky SH, Barnes BC. Machine learning transition temperatures from 2D structure. *J Mol Graph Model*. 2021;105, 107848.
30. Xie YH, Liu YJ, Hu RL, Lin X, Hu J, Pu XM. A property-oriented adaptive design framework for rapid discovery of energetic molecules based on small-scale labeled datasets. *RSC Adv*. 2021;11:25764–25776.
31. Fathollahi M, Sajady H. Prediction of density of energetic cocrystals based on QSPR modeling using artificial neural network. *Struct Chem*. 2018;29(4):1119–1128.
32. Fathollahi M, Sajady H. QSPR modeling of decomposition temperature of energetic cocrystals using artificial neural network. *J Therm Anal Calorim*. 2018;133: 1663–1672.
33. Yang CM, Chen J, Wang RW, Zhang M, Zhang CY, Liu J. Density prediction models for energetic compounds merely using molecular topology. *J Chem Inf Model*. 2021; 61(6):2582–2593.
34. Nguyen P, Loveland D, Kim JT, Karande P, Hiszpanski AM, Han TY. Predicting energetics materials' crystalline density from chemical structure by machine learning. *J Chem Inf Model*. 2021;61(5):2147–2158.
35. Ma ZL, Xu FL, Liu HY, Zhang WC. Predicting the detonating velocity of explosives based on artificial neural network and hybrid genetic algorithm. *Chin J Energetic Mater*. 2007;15(6):637–640.
36. Long XP, He B, Jiang XH, Wu X. Discussion on the VLW equation of state. *Chin J High Press Phys*. 2003;17(4):247–254.
37. Chandrasekaran N, Oommen C, Kumar VRS, Lukin AN, Abrukov VS, Anufrieva DA. Prediction of detonation velocity and N-O composition of high energy C-H-N-O explosives by means of artificial neural networks. *Propellants, Explos Pyrotech*. 2019; 44(5):579–587.
38. Wang F, Liu JH, Tian DY, et al. Prediction of the enthalpy of formation for aromatic polynitro compounds with artilicialneural network. *Chin J Energetic Mater*. 2004; 12(4):207–213.
39. Mathieu D. Molecular energies derived from deep learning: application to the prediction of formation enthalpies up to high energy compounds. *Mol Inf*. 2021, 2100064.
40. Keshavarz MH, Jaafari M. Investigation of the various structure parameters for predicting impact sensitivity of energetic molecules via artificial neural network. *Propellants, Explos Pyrotech*. 2006;31(3):216–225.
41. Kang P, Liu ZL, Abou-Rachid H, Guo H. Machine-learning assisted screening of energetic materials. *J Phys Chem A*. 2020;124(26):5341–5351.
42. Li G, Zhang CY. Review of the molecular and crystal correlations on sensitivities of energetic materials. *J Hazard Mater*. 2020;398(5), 122910.
43. Chen C, Liu DY, Deng SY, et al. Accurate machine learning models based on small dataset of energetic materials through spatial matrix featurization methods. *J Energy Chem*. 2021;63:364–375.
44. Huang XN, Li Cy,Tan KY, Wen YS, et al. Applying machine learning to balance performance and stability of high energy density materials. *iScience*. 2021;24(3), 102240.
45. Hou F, Ma Y, Hu Z, et al. Machine learning enabled quickly predicting of detonation properties of N-containing molecules for discovering new energetic materials. *Adv Theory Simul*. 2021;4(6), 2100057.
46. Song SW, Wang Y, Chen F, Yan M, Zhang QH. Machine learning-assisted high-throughput virtual screening for on-demand customization of advanced energetic materials. *Engineering*. 2022;10(3):99–109.
47. Song SW, Chen F, Wang Y, Wang KC, Yan M, Zhang QH. Accelerating the discovery of energetic melt-castable materials by a high-throughput virtual screening and experimental approach. *J Mater Chem*. 2021;9(38):21723–21731.

Xiao-lan Tian graduated from Sichuan University of Science & Engineering with a Bachelor of Engineering degree in 2019. She is currently pursuing master's degree in materials and chemical engineering in Southwest University of Science and Technology. Her research interest is the application of machine learning in energetic materials. E-mail: tismol@163.com

Yi Wang received his Ph.D. at Jilin University in 2016. From 2016 to 2018, he worked as postdoctoral fellow in Institute of Chemical Materials, China Academy of Engineering Physics (CAEP). Since 2018, he has worked as an associate professor at Institute of Chemical Materials, CAEP. His research interests mainly focus on design, synthesis, and characterization of new energetic materials. ORCID ID: https://orcid.org/0000-0003-1875-6597. E-mail: ywang0521@caep.cn

Xiu-juan Qi received her Ph.D. at University of Science and Technology of China in 2008 under the supervision of Prof. Qing-xiang Guo. Since 2015, she has been a member in Southwest University of Science and Technology. Her research interests mainly focus on the molecular design and property research of new energetic materials. E-mail: juanxiuqi@swust.edu.cn